





Colloque international Université Cheikh Anta Diop de Dakar

# Science ouverte au Sud

## ENJEUX ET PERSPECTIVES

## POUR UNE NOUVELLE DYNAMIQUE



23-25  
octobre  
2019

### Les entrepôts de données de la recherche. Exemple avec l'outil Dataverse

### Les entrepôts : panorama, critères de choix

Martine Barale, CIRAD-Dist



Avec le soutien de

En partenariat avec














## Un nouveau contexte : la science ouverte

**La science ouverte, c'est la diffusion sans entrave des publications et des données de la recherche pour :**

- Faire sortir la **recherche financée sur fonds publics** du cadre confiné des bases de données fermées pour augmenter son efficacité
- Construire un écosystème dans lequel la science est **plus cumulative, plus fortement étayée par des données, plus transparente, plus rapide et d'accès plus universel**
- Démocratiser l'**accès aux savoirs**, utile à la recherche, à la formation, à l'économie, à la société
- Favoriser les avancées scientifiques et l'**innovation**, dans les pays développés et dans les pays en développement.

La science ouverte constitue un **levier pour l'intégrité scientifique** et favorise la confiance des citoyens dans la science

## Les enjeux de l'accessibilité des données

### L'accessibilité des données devient un enjeu important de la recherche scientifique

- Permettre la reproduction et la vérification des résultats  
==> **"bonne science"**
- Eviter de recréer ce qui existe déjà  
==> **économie**
- Reprendre et approfondir les recherches précédentes  
==> **recherche plus cumulative**
- Alimenter des recherches dans **d'autres disciplines, d'autres contextes** ou à d'autres fins (enseignement...)
- Préserver et partager un **patrimoine**

Science ouverte au Sud, Dakar, 23-25/09/2019

3

## Partager ses données : entre injonctions...



Protocole de Nagoya

SPRINGER NATURE

WILEY

Incitation ou  
obligation  
de partage

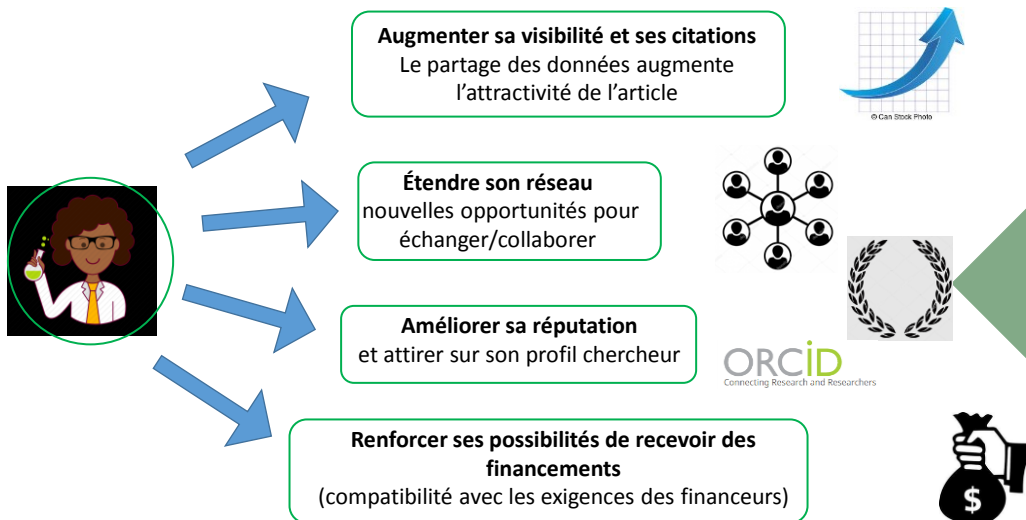
Reproductibilité  
Preuve  
Résultats accessibles à tout public  
Préservation  
Description et visibilité



Science ouverte au Sud, Dakar, 23-25/09/2019

4

## ... et bénéfices



5

## L'ouverture des données : oui mais...

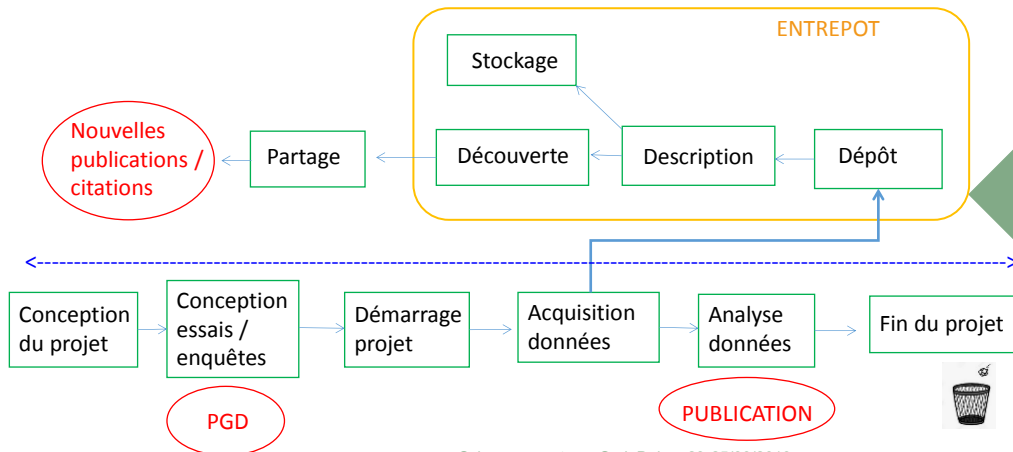
- **C'est compliqué ?...**
  - Utiliser une infrastructure de données
  - Rédiger un plan de gestion des données (DMP)
- **C'est cher ?...**
  - Les frais de stockage, d'organisation et si besoin de dépôt dans un entrepôt (la plupart sont gratuits) sont éligibles dans les appels à projets
- **C'est risqué ?...**
  - La présence de données sensibles (personnelles, secret défense, secret professionnel, secret industriel et commercial, risque pour la protection du potentiel scientifique...) constitue une exception aux exigences d'ouverture des données

▶ ▶ ▶ *"Aussi ouvert que possible, aussi fermé que nécessaire"*

Science ouverte au Sud, Dakar, 23-25/09/2019

6

## Changer les pratiques de gestion des données



7

## Les entrepôts de données : définition et typologie

Service en ligne permettant le dépôt, la description, la conservation, la recherche et la diffusion des jeux de données

### ➤ Editeur

- Données sous-jacentes aux publications

### ➤ Institution

- Données produites

### ➤ Thématique

- Données dans une discipline spécifique

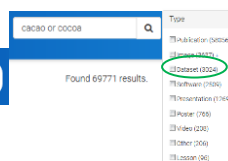
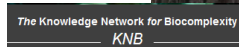
### ➤ Multidisciplinaire

- Données dans toutes les disciplines

### ➤ Mixte

- Publications et jeux de données liés ou non

### ➤ Public ou propriétaire, gratuit ou payant



Science ouverte au Sud, Dakar, 23-25/09/2019

8

## Les entrepôts de données : plus-value chercheur

### Augmentation de la visibilité des travaux

Les entrepôts sont scannés par des moteurs de recherche spécifiques



et moissonnés par des catalogues, intégrateurs, infrastructures de données... de plus en plus nombreux



►►► *Déposer ses données dans un entrepôt facilite leur découverte et leur réutilisation*

Science ouverte au Sud, Dakar, 23-25/09/2019

9

## Les entrepôts de données : fonctionnalités

- **Dépôt** et import des données
- Description des données (**métadonnées**)
- Attribution d'un identifiant pérenne (**DOI**)
- Stockage, archivage, **préservation**
- **Recherche** et accès aux données
- **Contrôle des accès** (embargo, licences...)
- Production de la **citation**
- +/-
- **Statistiques** d'usage (téléchargements)
- **Interopérabilité** avec d'autres plateformes ou outils
- Contrôle de **qualité** des métadonnées...

Science ouverte au Sud, Dakar, 23-25/09/2019

10

## Exemple 1 : Zenodo – ouvert à tous, gratuit



Répertoire généraliste de travaux de recherche et de données

<https://zenodo.org/>

Found 8250 results. Best match: June 24, 2016. Sort by: 850.

**Typical Sensor Defects Dataset**  
Ehlerbroker, Jan-Friedrich; Monks, Uwe; Lohweg, Volker

Thirteen datasets of sensor values, with one dataset without sensor defects (data\_standard.csv). All other datasets are based on the dataset without a defect, with the values of Temp\_Sensor\_2 modified to simulate different sensor defects:

- Sensor Drift: 1% per hour (data\_drift\_0\_001.csv), 2.5% per hour (data\_drift\_0\_0025.csv), 5% per hour (data\_drift\_0\_005.csv)
- Sensor Offset: 1°C Offset (data\_offset\_1.csv), 2°C Offset (data\_offset\_2.csv), 5°C Offset (data\_offset\_5.csv)
- Sensor Peaks: 1 Peak/Minute (data\_peak\_1.csv), 2 Peaks/Minute (data\_peak\_2.csv), 5 Peaks/Minute (data\_peak\_5.csv), 10 Peaks/Minute (data\_peak\_10.csv)
- Sensor Noise: 10 dB SNR (data\_noise\_10db.csv), 0 dB SNR (data\_noise\_0db.csv)

The datasets are given as comma-separated values in text files. The first column in each file holds time stamps, while the following columns hold the sensor values. The first entry in every column gives the name of the sensor. All datasets are zipped into one file (data.zip).

Additionally attached is configuration data (Configuration.pdf) for the sensor fusion approach that was used to classify the datasets.

For more information please contact the uploader.

47 views, 38 downloads. See more details.

Indexed in OpenAIRE

Publication date: June 24, 2016  
DOI: [10.5281/zenodo.54214](https://doi.org/10.5281/zenodo.54214)  
Keywords: [empty]  
License (for files): [Creative Commons Attribution 4.0 International](#)

Share: [Facebook] [Twitter] [LinkedIn] [Email] [Print]

Cite as

Type: Publication (4803), Dataset (274), Software (96), Poster (87), Image (38), Presentation (36), Other (12), Video (11), Lesson (3)

Files (4.1 MB):  
Name: Configuration.pdf, Size: 60.8 KB, Preview, Download  
Name: data.zip, Size: 4.0 MB, Preview, Download

Science ouverte au Sud, Dakar, 23-25/09/2019

11

## Exemple 2 : Dryad – payant (120 \$ jusqu'à 20 GB)

**DRYAD**

Explore Data | About | Help | Login

Explore and discover... Find the data you need

Search... Search [Q] Browse data publications

Find by...

Placename: North America, Europe, Australia, USA, Canada, United States, California, Africa, South America, More

Subject: Holocene, Adaptation, climate change, speciation, phylogeography, Population Genetics - Empirical, hybridization, Phenotypic Plasticity, phylogeny, More

Journal: Molecular Ecology, PLOS ONE, Evolution, Ecology and Evolution

Institutions: University of California System, University of Oxford, French

Find by location: [Map of the world with search bar]

Entrepôt généraliste mis en place par des revues et des sociétés savantes pour archiver les données qui sous-tendent les publications scientifiques.

Accueille aujourd'hui toutes sortes de données.

Login avec votre identifiant ORCID

<https://datadryad.org/>

Science ouverte au Sud, Dakar, 23-25/09/2019

12

## Exemple 3 : Harvard Dataverse

Entrepôt initié autour des sciences sociales mais qui accepte aujourd'hui des données dans toutes les disciplines.

<https://dataverse.harvard.edu/>

Basé sur le logiciel libre  
Dataverse, que toute  
institution peut utiliser

Science ouverte au Sud, Dakar, 23-25/09/2019

13

## Et en Afrique ?...

### ➤ Des entrepôts en **sciences humaines et sociales**



DataFirst is a research data service dedicated to giving open access to data from South Africa and other African countries. We promote high quality research by providing the essential Open Research Data infrastructure for discovering and accessing data and by developing skills among prospective users, particularly in South Africa. We undertake research on the quality and usability of national data and encourage data usage and data sharing.



#### Objectives of SADA are to

- acquire and catalogue survey data and related information,
- preserve such data against technological obsolescence and physical damage,
- re-disseminate such information for use by other researchers, for re-analysis of data, longitudinal and comparative studies, research training, teaching and policy-making decision purposes,
- formulate policies for the scope and content of data and data preservation,
- promote the optimal use of data.

### ➤ Des entrepôts dans le domaine de la **biodiversité**

#### WEST AFRICAN VEGETATION

a vegetation data network for West Africa

#### WEST AFRICAN VEGETATION - A VEGETATION DATA NETWORK FOR WEST AFRICA

This is a database for vegetation data from West Africa, i.e. phytosociological and dendrometric relevés as well as floristic inventories.



#### West African Plants

A PHOTO GUIDE



Science ouverte au Sud, Dakar, 23-25/09/2019

14

# Comment trouver le bon entrepôt ?

## De nombreux critères de choix

Caractéristiques	Fonctionnalités	Exigences
Pérennité de l'entrepôt	Préservation des données	Limite disciplinaire
Facilité de dépôt	Identifiant pérenne	Coûts
Facilité de recherche / découverte des données	Contrôle d'accès aux données	Types de données acceptés
Localisation du serveur	Traçabilité, provenance	Formats acceptés
Qualité de la description des données (utilisation de standards, contrôles,...)	Statistiques d'usage (consultation, téléchargement)	Licences possibles
	Versionning	Limite volumétrique
	Interopérabilité	
	Prépublication (accès des reviewers via lien privé)	

Adapté de : Esther Dzalé Yeumo, DIST INRA

Science ouverte au Sud, Dakar, 23-25/09/2019

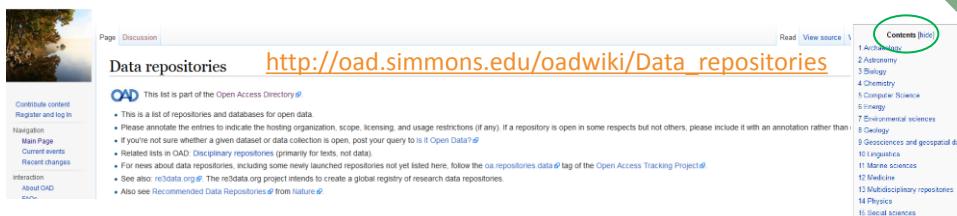
15

# Annuaire généralistes (tous types d'entrepôt)

## ➤ re3data



## ➤ Open access directory > data repositories



Science ouverte au Sud, Dakar, 23-25/09/2019

16



# Annuaire d'entrepôts par domaines

## Exemple dans les sciences biomédicales

**FAIRsharing.org**  
standards, databases, policies

Recherche d'entrepôts

Listes par type de données, sujet, espèce, pays, institution

<https://fairsharing.org/databases/>

Science ouverte au Sud, Dakar, 23-25/09/2019

17

# Annuaire mixtes (archives ouvertes, entrepôts)

## ➤ ROAR

### Registry of Open Access Repositories

Home About Search Search Content Browse  
Login | New Entry | Create Account

<http://roar.eprints.org/>

## ➤ OpenDOAR

**OpenDOAR**  
Browse Search Statistics Policy Tool Our APIs

Directory of Open Access Repositories  
OpenDOAR is a global directory of Open Access repositories and their policies

Recherche d'entrepôts

<https://v2.sherpa.ac.uk/opendoar/>

Science ouverte au Sud, Dakar, 23-25/09/2019

18

## Listes d'entrepôts recommandés (éditeurs, bailleurs)

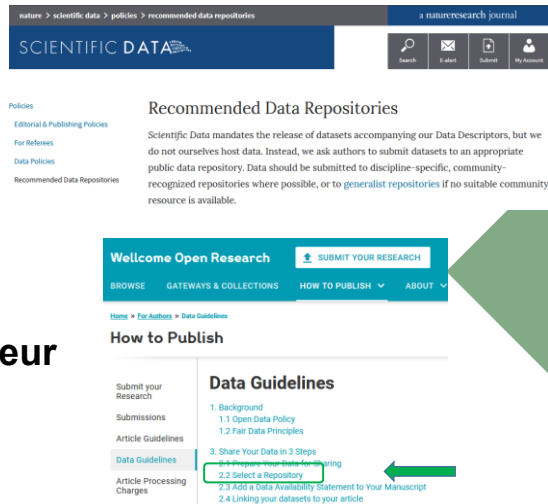
- **Si projet de publication**, vérifier la politique de la revue pour la diffusion des données
  - incitation / obligation
  - entrepôts préconisés

Expl : *Groupe Nature*

- **Si financement d'un bailleur**

idem

Expl : *Wellcome Trust*



Science ouverte au Sud, Dakar, 23-25/09/2019

19

## Les moteurs de recherche de données

- Permettent de rechercher des données dans **plusieurs entrepôts** en même temps, et proposent des filtres puissants
- **Données de tous types et dans tous les domaines** : environnement, sciences sociales, émanant d'agences gouvernementales, données de presse...
- **Moteurs spécialisés**
  - Datacite
  - DataSearch (Elsevier)
  - Google Dataset Search
  - Data Citation Index (accès payant)
- Un **moteur "mixte"** : BASE

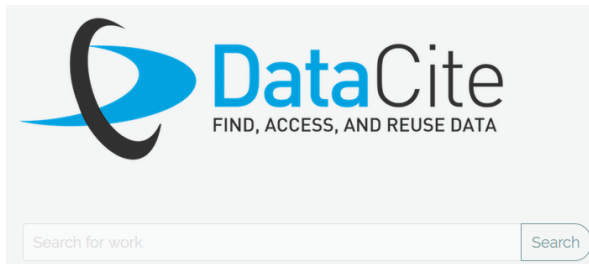


Science ouverte au Sud, Dakar, 23-25/09/2019

20

## Datacite

- Organisation internationale à **but non lucratif**
- Délivre des **DOI** aux institutions membres



<https://search.datacite.org/>

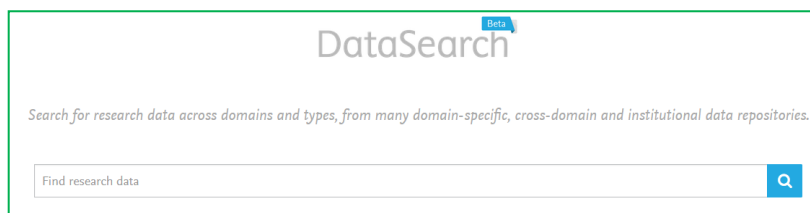
- Réunit les métadonnées et les liens vers toutes les ressources ayant un DOI Datacite dans une base librement interrogeable

Science ouverte au Sud, Dakar, 23-25/09/2019

21

## Elsevier DataSearch (Beta)

- Développé par un des principaux **éditeurs scientifiques**
- Indexe les contenus d'une dizaine d'entrepôts (dont Dryad, Pangaea, Harvard Dataverse) et les "supplementary data" dans ArXiv et Science Direct (revues Elsevier)



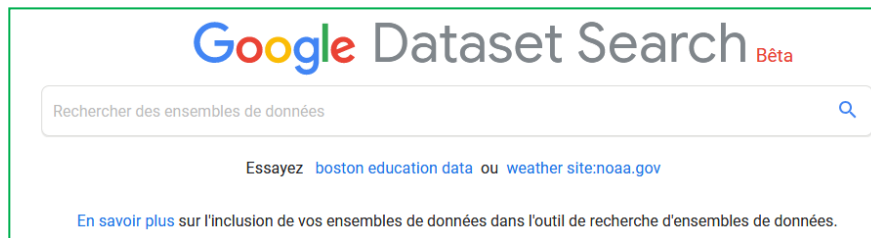
<https://datasearch.elsevier.com/>

Science ouverte au Sud, Dakar, 23-25/09/2019

22

## Google Dataset Search (Beta)

- Indexe "des milliers" d'entrepôts basés sur le standard schema.org (mais aucune liste n'est diffusée)
- Fonctionne sur le même modèle que Google Scholar



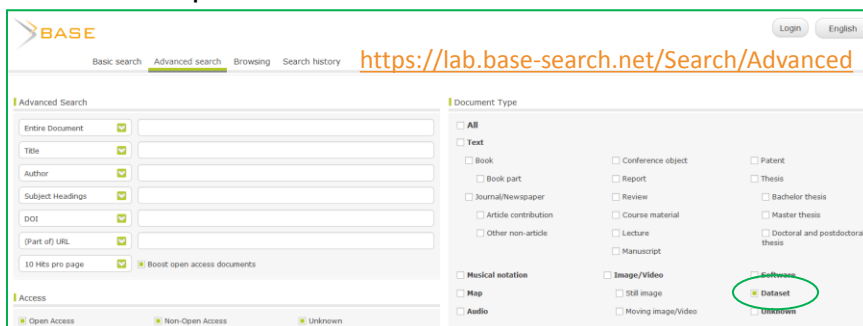
<https://toolbox.google.com/datasetsearch>

Science ouverte au Sud, Dakar, 23-25/09/2019

23

## BASE

- Moteur de recherche spécialisé dans les **ressources web académiques** : publications, données, cartes, images...
- Indexe toutes sortes de **ressources** (revues, entrepôts, collections) conformes au standard OAI-PMH
- A noter : permet de sélectionner les ressources en libre accès



Science ouverte au Sud, Dakar, 23-25/09/2019

24





Colloque international Université Cheikh Anta Diop de Dakar

# Science ouverte au Sud

## ENJEUX ET PERSPECTIVES

## POUR UNE NOUVELLE DYNAMIQUE



23-25  
octobre  
2019

Merci pour votre attention

## Des questions ?...

N'hésitez pas à nous contacter

[martine.barale@cirad.fr](mailto:martine.barale@cirad.fr)



Bureau de l'éducation  
au UNESCO

Avec le soutien de



En partenariat avec









